

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1.-20. (Cancelled).

21. (Currently amended) A system for clustering data comprising:
a computer executing a computer program performing at least the following:
receiving into the computer a plurality of data points for clustering;
receiving into the computer a size parameter for specifying the
number of data points to be moved—at one
timesimultaneously evaluated for inclusion in a cluster;
clustering the data points by using the size parameter to generate
clustered results;
determining whether the clustered results are satisfactory;
when the clustered results are satisfactory, stop clustering;
otherwise when the clustered results are not satisfactory, revise the
size parameter, perform clustering based on the revised size
parameter and the clustered results, and proceed to step
(d)determining whether the clustered results are satisfactory.

22. (Currently amended) The system as defined in claim 21 wherein
clustering the data points further comprises:
evaluating subsets of data points in each cluster for moving into every
other cluster by using a predetermined metric; wherein the number
of data points in the a subset is specified by the size parameter.

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

23. (Previously presented) The system as defined in claim 22 wherein evaluating subsets further comprises:

- determining a geometric center of the subset of data points being evaluated for a move;
- using the geometric center of the subset of data points and the predetermined metric to generate a value.

24. (Previously presented) The system as defined in claim 23 wherein evaluating subsets further comprises:

- determining whether the value is greater than zero;
- when the value is greater than zero, moving the subset of data points from a Move_From cluster to a Move_To cluster;
- when the value is not greater than zero, determining if there are more subsets to evaluate;
- when there are more subsets to evaluate, proceeding to evaluating the subsets;
- when there are no more subsets to evaluate, determining whether any point has moved;
- when a point has moved, proceeding to evaluating the subsets; and
- when no point has moved, stopping the processing.

25. (Currently amended) ~~The system as defined in claim 24 further comprising:~~ A system for clustering data comprising:

a computer executing a computer program performing at least the following:

- receiving into the computer a plurality of data points for clustering;
- receiving into the computer a size parameter for specifying the number of data points to be moved at one time;
- clustering the data points by using the size parameter to generate clustered results;
- determining whether the clustered results are satisfactory;

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

when the clustered results are satisfactory, stop clustering;
otherwise when the clustered results are not satisfactory, revise the size
parameter, perform clustering based on the revised size parameter
and the clustered results, and proceed to determining whether the
clustered results are satisfactory,

wherein clustering the data points further comprises:

evaluating subsets of data points in each cluster for moving into
every other cluster by using a predetermined metric, wherein
the number of data points in a subset is specified by the size
parameter,

wherein evaluating subsets further comprises:

determining a geometric center of the subset of data points
being evaluated for a move;

using the geometric center of the subset of data points and
the predetermined metric to generate a value;

determining whether the value is greater than zero;

when the value is greater than zero, moving the subset of
data points from a Move_From cluster to a
Move_To cluster;

when the value is not greater than zero, determining if there
are more subsets to evaluate;

when there are more subsets to evaluate, proceeding to
evaluating the subsets;

when there are no more subsets to evaluate, determining
whether any point has moved;

when a point has moved, proceeding to evaluating the
subsets; and

when no point has moved, stopping the processing,

wherein each data point has a membership with one cluster; and

wherein moving the subset of data points from a Move_From cluster to a
Move_To cluster further comprises simultaneously updating the

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

membership of at least two data points from the membership of the Move_From cluster to the membership of the Move_To cluster.

26. (Previously presented) The system as defined in claim 24 wherein moving the subset of data points from a Move_From cluster to a Move_To cluster further comprises:

updating the count of the Move_From cluster;
 updating the center of the Move_From cluster;
 updating the count of the Move_To cluster;
 updating the center of the Move_To cluster.

27. (Previously presented) The system as defined in claim 21 wherein revising the size parameter further comprises decreasing the size parameter.

28. (Cancelled).

29. (Currently amended) ~~The computer-readable medium of~~ The system as defined in claim 28-22 wherein the predetermined metric of the computer program comprises the following expression:

$$\frac{\frac{n_i}{n_i - |U|} |m_u - m_i|^2 - \frac{n_i}{n_i + |U|} |m_u - m_i|^2}{\frac{n_i}{n_i - |U|} |m_u - m_i|^2 - \frac{n_i}{n_i + |U|} |m_u - m_i|^2}$$

where U is the subset of data points being evaluated for the move, $|U|$ is the size of U that is specified by the size parameter, m_u is the geometric center of U , m_i and m_j are the centers of the clusters and n_i and n_j are the counts of the clusters.

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

30. (Previously presented) The system as defined in claim 21 wherein the system is utilized in one of a data mining application, customer segmentation application, document categorization application, scientific data analysis application, data compression application, vector quantization application, and image processing application.

31. (New) The system as defined in claim 21 wherein determining whether the clustered results are satisfactory further comprises:

determining whether a change in a performance function is less than a predetermined value.

32. (New) The system as defined in claim 24, further comprising:

wherein each data point has a membership with one cluster; and

wherein moving the subset of data points from a Move_From cluster to a Move_To cluster further comprises simultaneously updating the membership of at least two data points from the membership of the Move_From cluster to the membership of the Move_To cluster.

33. (New) A computer-implemented method for clustering data points, comprising:

receiving a plurality of data points in a computer system;

partitioning the plurality of data points into a plurality of clusters wherein each data point is a member of one cluster of the plurality of clusters;

evaluating a plurality of data points in a first cluster of the plurality of clusters simultaneously to determine whether the plurality of data points in the first cluster should be moved to a second cluster of the plurality of clusters; and

moving the plurality of data points simultaneously from the first cluster to the second cluster if the determination is that the plurality of data points should be moved.

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

34. (New) The computer-implemented method of claim 33, further comprising:
receiving a size parameter that specifies a number of data points, wherein
the plurality of data points in a first cluster comprises the number of
data points specified by the size parameter.

35. (New) The computer-implemented method of claim 34, wherein evaluating
a plurality of data points simultaneously further comprises applying a
predetermined metric in the evaluation, wherein the predetermined metric uses
the size parameter.

36. (New) The computer-implemented method of claim 35, wherein the
predetermined metric comprises the following expression:

$$\frac{n_i}{n_i - |U|} |m_u - m_i|^2 - \frac{n_i}{n_j + |U|} |m_u - m_j|^2$$

wherein U is the plurality of data points being evaluated, $|U|$ is a size of U that is
specified by the size parameter, m_u is a geometric center of U , m_i and m_j are
geometric centers of the first cluster and the second cluster, and n_i and n_j are
counts of the first cluster and the second cluster.

37. (New) A system for clustering data points comprising:
a computer that stores a plurality of data points for clustering;
means for partitioning the plurality of data points into a plurality of clusters;
means for evaluating whether an aggregated move of a subset of data
points in a first cluster of the plurality of clusters to a second cluster
of the plurality of clusters improves the partitioning; and
means for performing the aggregated move of the subset of data points to
the second cluster responsive to the results of the evaluation
means.

38. (New) The system of claim 37, wherein the means for evaluating utilizes a
predetermined metric to simultaneously evaluate the data points in the subset.

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

39. (New) The system of claim 38, wherein the predetermined metric comprises the expression:

$$\frac{n_i}{n_i + |U|} |m_u - m_i|^2 - \frac{n_j}{n_j + |U|} |m_u - m_j|^2$$

wherein U is the subset of data points being evaluated, $|U|$ is a size of U that is specified by a size parameter, m_u is a geometric center of U , m_i and m_j are geometric centers of the first cluster and the second cluster, and n_i and n_j are counts of the first cluster and the second cluster.

40. (New) The system of claim 37, wherein the means for evaluating receives a size parameter that determines the number of data points included in the subset.

41. (New) The system of claim 37, wherein the means for evaluating includes means for generating a geometric center of the subset of data points based on a first plurality of data points in the first cluster and a second plurality of data points in the second cluster.

42. (New) The system of claim 37, wherein the means for performing an aggregated move includes means to simultaneously move the data points in the subset of data points from the first cluster to the second cluster.

43. (New) The system of claim 37, wherein the means for performing an aggregated move includes means to determine a geometric center of the first cluster and a geometric center of the second cluster after the subset of data points is moved, wherein a geometric center of the subset of data points is used in the determination.

44. (New) A computer-implemented method for clustering data points comprising:

receiving in a computer system a plurality of data points for clustering;

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

partitioning the plurality of data points into a plurality of clusters; and
repartitioning the plurality of data points among the plurality of clusters,
wherein repartitioning comprises:

using a predetermined metric to evaluate subsets of data points in
each cluster of the plurality of clusters for moving into every
other cluster of the plurality of clusters, wherein the number
of data points in each subset is determined by a size
parameter; and

moving data points in a subset of data points simultaneously from a
first cluster of the plurality of clusters to a second cluster of
the plurality of clusters if the evaluation of the subset
determines that the subset should be moved into the second
cluster.

45. (New) The computer-implemented method of claim 44, further comprising
determining whether the repartitioning is satisfactory;
if the repartitioning is not satisfactory, performing:
changing the size parameter;
repartitioning the plurality of data points; and
determining whether the repartitioning is satisfactory
until the repartitioning is satisfactory.
46. (New) The computer-implemented method of claim 45, wherein
determining whether the repartitioning is satisfactory comprises determining
whether at least one data point was moved during the repartitioning.
47. (New) The computer-implemented method of claim 45, wherein
determining whether the repartitioning is satisfactory comprises determining
whether a change in a performance function is less than a predetermined value.

Appl. No. 09/684,488
Amdt. dated July 12, 2004
Reply to Office action of May 17, 2004

48. (New) The computer-implemented method of claim 44, wherein moving data points in a subset of data points simultaneously further comprises:

updating counts of the first cluster and the second cluster; and
recalculating geometric centers of the first cluster and the second cluster.

49. (New) The computer-implemented method of claim 44, wherein the predetermined metric is computed using at least the size parameter, geometric centers of the subset being evaluated, the first cluster, and the second cluster, and counts of the first cluster and the second cluster.